



Copyright © 1997 by Douglas Adams

*"I'm going to put him back to bed in a minute, Carol.
I just need Jeffrey to check my HTML tags for errors..."*

Part 1:

HTML 4.0 Fundamentals

Creating Hypertext Web Pages

Part 1 Description

Welcome to *HTML 4.0 Fundamentals*, the first section in DDC's *Mastering HTML* series. This course is designed as a thorough and detailed introduction to the HyperText Markup Language—more commonly known as HTML. This course covers elements of HTML Specifications 3.2 and 4.0 as maintained by the World Wide Web Consortium.

Part 1 Objectives

This course is intended for people who want to learn to use HTML to create Web pages suitable for publishing via the Internet and viewing with a Web browser, such as Netscape Navigator/Communicator or Microsoft Internet Explorer.

- Hypertext and Hypermedia
- Hyperlinks (HTML anchors)
- HTML tag syntax and usage
- Tag attributes
- Lists (ordered and unordered)
- Character references (entity and numeric)
- Inline images
- Background colors, background patterns, and text colors
- <META> tag

Part 1 Setup

DDC's *HTML 4.0 Fundamentals* requires minimal PC configuration and setup. Three primary elements are required for this course: Web browser, text editor, and student files.

During this entire course, you (or your students) will toggle back and forth between your Web browser and your text editor as you create Web pages following the exercises in this course.

Required Software

Two software applications are required to take or teach this course:

- A Web browser (Netscape Navigator 4.0x or higher or Microsoft Internet Explorer (IE) 5.0x or higher recommended). Navigator can be downloaded from the Netscape Web site at www.netscape.com; Microsoft IE can be downloaded from Microsoft's Web site at www.microsoft.com/ie.
- A text editor (Microsoft Windows Notepad recommended). Notepad can be launched in Windows 95/98/2000 and Windows NT from the Start button on the task bar (choose **Programs** ► **Accessories** ► **Notepad**).

1

Lesson 1 **What is HTML?**

Lesson Topics

- ▶ Hypertext
- ▶ The HyperText Markup Language
- ▶ What is an HTML Document?
- ▶ HTML Document Structure
- ▶ HTML Document Characteristics
- ▶ Lesson 1 Summary

Hypertext

Hypertext is the method by which documents, files, and multimedia data are organized (by publishers), referenced, and navigated (by users) via the Internet's World Wide Web. The concept of hypertext can be traced back to the ideas of Dr. Vannevar Bush, science advisor to President Roosevelt.¹ The theory of hypertext was expanded during the 1960s and 1980s by Ted Nelson. Nelson coined the terms "hypertext" and "hypermedia" in 1965.

There are three primary objectives of hypertext information systems:

- to make available extremely large volumes of information from a wide variety of sources via a single medium (the World Wide Web);
- to make this body of information easy to navigate and to make specific data easy to locate;
- to allow you to explore and navigate this information via an arbitrary path of your choice.

Hyperlinks

A *hyperlink*, or link, is a visible navigational anchor in a hypertext document that refers (points) to another hypertext document or multimedia data object (audio file, video file). A hyperlink can refer to a different location in the same hypertext document or to a specific location in a different hypertext document. There is no limit to the number of hyperlinks that can be placed within an individual hypertext document, but each hyperlink can only point to a single document or multimedia data object.

Hypertext Publishing vs. Conventional Publishing

In conventional publishing, an author attempts to eliminate extraneous, supplemental, or "peripheral" information from a document via the editing process. In hypertext publishing, however, peripheral information need not be excluded; it may simply be relegated to a separate hypertext document or collection of documents that are accessible via hyperlinks.

Hypertext publishing creates an environment in which each user determines what information is central and important and what is peripheral or of minimal value. Some contemporary literary experts, such as Dr. George Landow of Brown University, believe that hypertext enables the user to play a part as an author. Although the user does not write the information presented, he or she is able to navigate his or her own path through a body of hypertext documents, viewing only that information that is of interest or subjective value.

¹ July 1945 *Atlantic Monthly* article "As We May Think."

Nonlinear Information

When reading a book or magazine article, information is displayed in a linear sequence and only makes sense if it is consumed in a specific, predefined order (progressing from page 1 to page 2 to page 3, and so forth).

If consumed in any other order, linear publications lose their meaning. By contrast, hypertext systems not only *allow* you to navigate through a hypermedia “space” following the path of your choice, but *require* you to do so to consume the available information.

The basis of hypertext is *nonlinear*, nonhierarchical, fluid, and “borderless” information. True hypertext publications do not offer or suggest any predefined path through a body of information. Modern tools for navigating hypertext information systems allow you to quickly revisit a previously viewed document in order to backtrack and, if you choose, select a different hyperlink available in that document.

Conventional vs. Nonlinear Publications

Many, but not all, paper-based publications also assume that the reader will consume the entire body of the text, as in the case of a novel, magazine article, or newspaper article. Reference materials, such as a dictionary, thesaurus, or encyclopedia, however, are designed to provide small pieces of valuable information quickly. Such publications are suitable for presentation in a hypertext format and navigated via hyperlinks.



“Many, perhaps most, writers have been frustrated by the problem of choosing a sequence for the ideas they are presenting. Any sequence is generally arbitrary, and what is right for one reader may be wrong for another.”

— Ted Nelson, *inventor of hypertext and hypermedia*, 1982

Hypermedia

Hypermedia is the fusion of hypertext with multimedia files, such as audio, video, and animation. Hypermedia allows Web documents to include hyperlinks that can connect you to a variety of multimedia files, as well as specialized multimedia data, such as virtual reality environments and interactive applications.

A problem many users experience with hypermedia is the amount of time it takes to download multimedia files. Most Internet connections have technical limitations and relatively poor data transfer capabilities. Several new multimedia data formats for the Web have been developed to overcome these technical limitations. The most promising category of multimedia data is *streamed data*. Streaming applies to audio and video formats and makes these otherwise impractical multimedia formats consumable by the average user.

The HyperText Markup Language

The *HyperText Markup Language*, or HTML, is a network-based document scripting language that allows multimedia hypertext documents and related multimedia data to be authored, edited, published, referenced, and navigated via the World Wide Web.

HTML comprises four primary elements, or concepts:

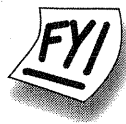
- hypertext and hypermedia computer network-based publishing;
- a scripting language (HTML tag family);
- a software application to locate, download, interpret, and display Web data (Web browser, such as Netscape Navigator or Microsoft Internet Explorer);
- a Web-specific data transfer protocol (HyperText Transfer Protocol or HTTP).

Scripting Language

HTML is a *scripting language*.² It is a language for “scripting” documents (analogous to a playwright authoring the “script” of a play) that utilizes a family of syntactical elements called *tags*. These tags are interpreted by a Web browser and tell the browser what to display (content) and how to display it (format).

Thus, an HTML document (commonly called a Web page) is comprised of two discrete elements:

1. content (the words and images in the page); and
2. instructions (HTML tags that command a Web browser to display the content in a particular manner).



A scripting language is very different from a programming language. Programming languages—such as C++, Pascal, or Java—involve both application files and associated data files. HTML, on the contrary, is self contained; the instructions (HTML tags) and data (non-tag HTML document content) are contained within a single file (the HTML document).

² A scripting language is sometimes referred to as a *markup language*.

The HTML Specification

The official technical rules of HTML—including definitions and usage parameters of the complete tag set, tag attribute set, and character reference entities—are called the *HTML Specification*³. Since its introduction in 1992, the HTML Specification has evolved from version 1.0, to version 2.0, to version 3.2, to the current version 4.0.

The HTML Specification is maintained by the World Wide Web Consortium or W3C. The W3C, based in Cambridge, Massachusetts, has been instrumental in establishing and maintaining technical standards for the World Wide Web since 1994.



“Insisting on HTML 4.0 compliance now will preserve your free choice of suppliers of Web software, tools and applications well into the future. With HTML 4.0, any Web application can be vendor independent. There really is no excuse for tying yourselves or your partners to proprietary solutions.”

— Tim Berners-Lee, *Director of the W3C, 1998*

HTML Extensions

Although the W3C publishes the official HTML Specification, both Netscape Communications Corporation and Microsoft Corporation—the manufacturers of the two leading Web browsers—frequently publish “extensions” to the HTML Specification. *Extensions* are tags, tag attributes, and character references proposed or published by any party or organization other than the W3C (typically a browser developer/manufacture).

HTML extensions, by definition, are unofficial additions to the HTML Specification. Even so, many HTML extensions are extremely popular and are in use on millions of Web sites. However, because they deviate from the official HTML standard, these extensions create a lack in conformity and consistent interpretation of Web pages among available Web browsers.

For example, assume Microsoft Corporation introduces a new HTML tag that only the Microsoft Internet Explorer Web browser recognizes and can properly interpret. HTML scripters and webmasters who use this proprietary tag risk all non-Microsoft browsers misinterpreting and not properly displaying pages containing this unofficial tag. You should avoid proprietary HTML extensions unless they are recognized by both the current and the previous versions of *both* Netscape Navigator/Communicator and Microsoft Internet Explorer.

³ The complete HTML 4.01 Specification can be downloaded from the W3C Web site at www.w3.org.

Web Browsers

Web browsers are software applications that run on client computers. In simplified terms, browsers do two things:

- interpret HTML documents;
- display HTML documents.

Despite the fact that HTML is a cross-platform, universal scripting language, browsers from different developers (Netscape and Microsoft) interpret and display HTML documents slightly differently. Sometimes the *same* version of the *same* browser on a different computer operating system (Windows, Macintosh, UNIX) will display a single HTML document differently.



“As computer CRT screens become more and more available, there is less and less reason for printing on paper. The costs of wood pulp and gasoline, the long lead times of editorship and production, the increasing divergence of specialized interests, and the lowering cost of computers with screens, of disk storage, and digital communications, all suggest this.”

— Ted Nelson, *inventor of hypertext and hypermedia*, 1982

Discrepancies in HTML document interpretation and display are also apparent between different versions of the same browser for the same computer platform. As a result of the rapid evolution of the HTML Specification (as promoted by the W3C, Netscape, and Microsoft), later versions of any browser provide considerably more functionality and capability than earlier versions. Thus, HTML documents may be interpreted properly in the *latest* version of a browser but misinterpreted and improperly displayed in the *previous* version of the same browser for the same computer platform.



To ensure the most universal and correct interpretation and display of the pages you create among all users of the Web, always test your HTML documents by opening them in both the current and the previous versions of Netscape Navigator and Microsoft Internet Explorer.

While this testing is a time consuming process and requires you to have four Web browsers installed on your computer, it helps ensure the ability of as many Web users as possible to properly interpret your Web pages, saving you embarrassment and lost revenues.

HyperText Transfer Protocol

The *HyperText Transfer Protocol*, or HTTP, is a data communications protocol that operates in conjunction with the Internet's universal TCP/IP protocol. Technically speaking, HTTP is a sub-protocol of TCP/IP.

HTTP makes it possible for a single HTML document to be archived on any type of server computer and viewed using any type of client computer (Windows, DOS, Macintosh, UNIX) and any Web browser (Netscape Navigator, Netscape Communicator, Microsoft Internet Explorer, Opera). HTTP is the technical foundation that provides the true cross-platform capability of HTML. Basically, HTTP allows Web servers and Web browsers to communicate with one another.

Transfer Protocol

A *transfer protocol* is the method by which information is transferred between a server computer and a client computer across the Internet. The transfer protocol you specify in a link in an HTML document determines the type of server to which you are referring, be it a Web server, an FTP server, a news server (newsgroups, also called Usenet), or a mail server (e-mail).

The major transfer protocols used in HTML are listed in Table 1-1.

Transfer Protocol	Server Type	URL Syntax
FTP	FTP	ftp://ftp.domain.com
HTTP	Web	http://www.domain.com
MAILTO ⁴	E-mail	mailto:username@domain.com
NEWS	Newsgroup (Usenet)	news://news.domain.com nntp://news.domain.com

Table 1-1: Major transfer protocols supported by Web browsers

⁴ MAILTO is not a true transfer protocol, but instead a URL prefix that instructs a browser to automatically launch a user's default e-mail application and open a pre-addressed outgoing message, as you will learn during this course. Also note that MAILTO, unlike the other true protocols, lacks forward slashes.

What is an HTML Document?

HTML documents are files archived on Web servers that can be accessed by Internet users via Web browsers (Netscape Navigator, Microsoft Internet Explorer).

HTML documents:

- must be in ASCII file format⁵;
- may be any byte count size;
- may assume either portrait, landscape, a combination of both, or other non-traditional layout schemes;
- may include two chief elements: 1) HTML script (tags, tag attributes, and character entities) and 2) non-HTML content (text, images, or multimedia data).

Information other than HTML script or non-HTML content should be avoided.

Creating HTML Documents

HTML documents can be created using any standard text editor, word processor, or HTML editor⁶ (Microsoft FrontPage 2000⁷, HotDog Pro, HoTMetaL PRO). The format of an HTML document must be ASCII text (sometimes called *plain text* or *DOS text*).

HTML documents not formatted in ASCII text will be partially or totally misinterpreted by a Web browser. Thus, before previewing HTML documents in a Web browser or uploading them to a Web server, remember to format them in ASCII format.



When creating an HTML document with a word processor, be sure to save the file in ASCII format.

⁵ ASCII format is also called *plain text* format or *DOS text* format.

⁶ For more information regarding HTML editors, see Appendix F: *HTML Editors*.

⁷ See DDC's *Mastering FrontPage 2000 Series* for more information regarding Microsoft FrontPage.

HTML Document Structure

Traditional desktop publishing and journalistic design and layout dictate three sections to a physical page: 1) a header; 2) a body; and 3) a footer. In Web publishing with HTML, however, only two sections make up a document.

A properly formatted HTML document contains two sections:

- Head section
- Body section

Head Section

The Head section of an HTML document provides browsers with specific logical data pertaining to the Web page, such as the document title. None of the information provided by an HTML author in the Head section of an HTML document appears in the viewing area of the Web browser.

Body Section

The Body section of an HTML document contains all of the information that is displayed in the browser. While some logical instructions are available that are scripted outside of the Body section in an HTML document, none of this information is visible by a user in the content area of the browser.



“The overarching vision I propose, then, we might call a ‘hyperworld’—a vast new realm of published text and graphics, all available instantly; a grand library that anybody can store anything in...with links...available as options to anyone who wishes to publish them.”

— *Ted Nelson, inventor of hypertext and hypermedia, 1982*

HTML Document Characteristics

An HTML document must be properly formatted in order for any Web browser to interpret it correctly. To ensure correct formatting, HTML authors must understand the various characteristics of HTML documents.

White Space

Web browsers recognize no more than a single space between “content” words in an HTML document. Web browsers also ignore the following elements:

- line breaks;
- paragraph breaks;
- multiple non-breaking horizontal spaces.

The HTML script excerpt below, despite its formatting, produces the result shown in Figure 1-1. To force text to wrap or insert white space, specific tags must be used. The excerpt does not contain these tags, and thus most of the additional spaces and line breaks are disregarded in the browser. The Break tag and the Paragraph tag are among the tags that serve this purpose, both of which are covered in Lesson 2: *HTML Tags*.

This is an example of standard u n f o r m a t t e d b o d y t e x t d i s p l a y e d
within
an H T M L d o c u m e n t . I n t h e u p c o m i n g e x e r c i s e s , y o u w i l l
learn to use
Heading Level t a g s t o s t r u c t u r e y o u r W e b p a g e s a n d f o r m a t t i n g
t a g s t o f o r m a t t h e t e x t i n y o u r W e b p a g e s .

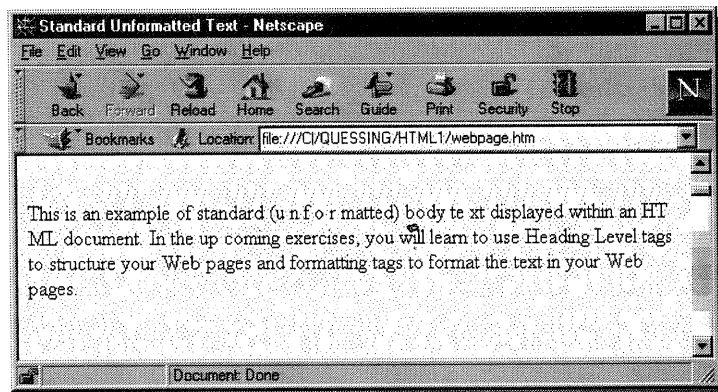


Figure 1-1: Result of poorly formatted script

Lesson 1 Summary

- ▶ HTML is a scripting language (sometimes called a *markup language*). Scripting languages differ substantially from programming languages.
- ▶ The official HTML Specification is maintained by the World Wide Web Consortium (W3C). The current version of the Specification is 4.0 and can be downloaded in a variety of file formats from the W3C's Web site at www.w3.org.
- ▶ HTML encompasses four primary technical elements: 1) hypertext and hypermedia publishing; 2) a scripting language; 3) a client software application to download, interpret, and display HTML documents; and 4) a Web-specific data transfer protocol (HTTP).
- ▶ HTML is a cross-platform scripting language. A document formatted in HTML can be viewed using any Web browser on any computer platform (DOS, Macintosh, UNIX, or Windows).
- ▶ Uniform Resource Locators (URLs) are the addressing scheme of Web pages and Web sites, and are a syntactical and technical conglomeration of the Web transfer protocol (HTTP), the server address, the directory path, and the file name. Some servers are configured to require only the transfer protocol and server name (<http://www.domain.com>) in order to access a Web page.
- ▶ Web browsers support several Internet transfer protocols. In addition to the HTTP protocol for HTML documents and related multimedia data available via the Web, popular browsers also support the FTP, MAILTO, and NEWS transfer protocols.
- ▶ An HTML document has two sections: a Head and a Body. The Head section traditionally contains only the title of the individual HTML document. The Body section contains the portion of the HTML document not contained in the Head section.
- ▶ HTML editors are valuable for those with a knowledge of HTML. They also help HTML authors to: 1) expedite the creation and editing of HTML documents, and 2) ensure the accuracy and validity of HTML script.
- ▶ Web browsers download, interpret, and display HTML documents. In sequence, browsers: 1) locate HTML documents; 2) download HTML documents; 3) interpret HTML documents; and 4) display Web pages.